

GIS Technologie News

SOGI Informationsblatt 1-2013

Haben 80% aller Informationen einen Raumbezug?

Einführung

Die These, dass 80% aller Informationen einen Raumbezug haben ist in der GIS-Welt weit verbreitet. Sie entstand parallel zum Aufkommens der ersten Geoinformationssysteme in 1980er Jahren und dient seitdem im gesamten Spektrum der Branche als Marketinginstrument. Jedoch findet sich bei der Recherche nach dem Ursprung der These keine Studie, die die Zahl empirisch be- oder auch widerlegt. Die frühesten Quellen zeigen, dass die These ursprünglich auf kommunale Datenbestände bezogen war, da diese viele Referenzen zu Adresdaten, Koordinaten und räumlichen Identifikationsnummern (z.B. Flurstücksnummern, Postleitzahlen) enthalten. Inzwischen wurde die Annahme jedoch (über)generalisiert und findet sich auch in allgemeinen Kontexten, nicht selten auch in wissenschaftlichen Publikationen. Eine aktuelle empirische Untersuchung am Lehrstuhl für Kartographische Kommunikation leistet einen Beitrag dazu, die identifizierte Forschungslücke zu schließen.

Wikipedia als Korpus und Netzwerk

Für die möglichst generische Untersuchung der These wurde ein Korpus gesucht, welcher eine Vielzahl an Informationen beinhaltet, ohne bestimmte Fachgebiete zu über- oder unterrepräsentieren. Hierfür eignet sich die Onlineenzyklopädie Wikipedia. Die Wikipedia kann dabei als Netzwerk modelliert werden. In einem gerichteten Graphen ergeben die Links zwischen den einzelnen Artikeln die Kanten und die Artikel selbst sind die Knoten. Aktuell enthalten ungefähr 18% der Artikel der deutschen Wikipedia, wie z.B. Städte, Berge oder Flüsse, eine explizite Kennzeichnung mit geographischen Koordinaten. Solche Artikel können als georäumliche Artikel bezeichnet werden. In unserem Ansatz werden analog zu sozialen Netzwerken, bei denen Kontakte ersten, zweiten, ..., Grades unterschieden werden, Artikel mit Raumbezug ersten, zweiten, ..., Grades berechnet. Dies geschieht, in dem für jeden Artikel im Netzwerk der jeweils kürzeste Weg zum nächstgelegenen georäumlichen Artikel bestimmt wird.

Kognitive Validierung

Da die eingefügten expliziten Koordinaten und die Verlinkungen der Artikel bereits einen kognitiven Prozess widerspiegeln, kann der analytisch bestimmbare Netzwerkgrad bereits als Ergebnis eines kognitiven Prozesses angesehen werden. In einer empirischen Untersuchung wurde darüber hinaus untersucht, inwiefern der Netzwerkgrad des Raumbezuges mit der Einschätzung bei der Befragung von Probanden korreliert. Dabei ging es unter anderem um die Frage, bei welchem Grad des Raumbezuges die menschliche Wahrnehmung zwischen raumbezogener Information und nicht-raumbezogener Information trennt. Hierfür wurde eine Stichprobe von 2.200 Wikipediaartikeln genutzt. Die Studienteilnehmer bekamen die Aufgabe, jeden Artikel anhand des Titels einer der vorgegebenen und häufig genutzten Kategorien ‚Direkter Raumbezug‘, ‚Indirekter Raumbezug‘ und ‚Kein Raumbezug‘ zuzuordnen.

Ergebnisse

Aufgrund der starken Vernetzung der Wikipediaartikel ergab sich für den Netzwerkansatz, dass es im wesentlichen nur vier Grade des Raumbezuges gibt: Grad 0 tritt zu 17,6% auf, Grad 1 zu 60,9%, Grad 2 zu 20,7% und Grad 3 zu 0,8%. Die drei relevantesten Grade des Raumbezuges – Grad 0 bis Grad 2 – wurden genauer untersucht. Prototypische Vertreter für Grad 0 sind z.B. ‚Zürich‘, ‚Fränkische Alb‘ und ‚Teldafax‘, für Grad 1 seien ‚Friedrich Schiller‘, ‚Französische Revolution‘ und ‚Aluminium‘ genannt, für Grad 2 sind abstrakte Begriffe, wie ‚lineare Funktion‘, ‚Synthie Pop‘ oder ‚Schießpulver‘, typisch. Im Netzwerkansatz hat somit jeder Artikel in Wikipedia einen (gestuften) Raumbezug. Somit haben hier gewissermaßen 100% aller Informationen einen Raumbezug. Die Befragung der Probanden diente nun der Überprüfung, für welchen Anteil an Informationen von Menschen der Raumbezug als relevant eingeschätzt wird. Es zeigte sich, dass Grad 0 Artikel zu 90% der Kategorie ‚direkter Raumbezug‘ zugeordnet wurden, Grad 1 Artikel

wurden etwa zu gleichen Teilen den Kategorien ‚indirekter Raumbezug‘ (39%) und ‚kein Raumbezug‘ (44%) zugeordnet und Grad 2 Artikel wurden mehrheitlich als ‚kein Raumbezug‘ (72%) kategorisiert. Aus der Kombination beider Ansätze lässt sich ableiten, dass in der deutschen Wikipedia 57% aller Artikel als ‚direkt‘ (27%) oder ‚indirekt‘ (30%) raumbezogen kategorisiert werden und 43% als ‚nicht-raumbezogen‘. Aufgerundet müsste die 80%-These demnach als ‚60%-These‘ neu formuliert werden.

Diskussion

Die verwendete Methodik hat einige Einschränkungen. Beispielhaft sei hier erwähnt, dass Wikipediaartikel häufig auf der Ebene von Konzepten verfasst sind. So wird beispielsweise das Konzept Haus hinsichtlich des Raumbezuges sicher anders beurteilt, als die konkrete Instanz ‚Weißes Haus (Washington D.C.)‘. Weiterhin ist die Wikipedia eine Enzyklopädie und weicht damit von der ursprünglichen Herkunft der 80%-These – kommunale Datenbestände – ab. Die Ergebnisse sollten somit nicht übergeneralisiert werden.

Publikationen

Für eine ausführliche Darstellung der Ergebnisse sei verwiesen auf:

- Hahmann, S.; Burghardt, D. (2012): Forschungsergebnisse zur Frage: Haben 80% aller Informationen einen Raumbezug? In: gis.SCIENCE - Die Zeitschrift für Geoinformatik, Nr. 3, 25, Wichmann, S.101–108
- Hahmann, S.; Burghardt, D.: How much information is geospatially referenced? Networks and Cognition. In: International Journal of Geographical Information Science, im Druck, DOI: 10.1080/13658816.2012.743664
- Pre-Prints beider Publikationen sind verfügbar unter:
 - http://kartographie.geo.tu-dresden.de/geospatial_experiment/Hahmann_Burghardt_gisscience_Preprint.pdf
 - http://kartographie.geo.tu-dresden.de/geospatial_experiment/Hahmann_Burghardt_IJGIS_Preprint.pdf

Fachgruppe GIS Technologie

technologie@sog.ch

Stefan Hahmann (wissenschaftlicher Mitarbeiter und Doktorand)

Dirk Burghardt (Professor für Kartographische Kommunikation)

Beide Autoren arbeiten derzeit am Institut für Kartographie der TU Dresden.